Algorithms Seminar 1999–2000,
F. Chyzak (ed.), INRIA, (2000), pp. 27–30.

Available online at the URL
http://algo.inria.fr/seminars/.

Random Group Automata

Cyril Nicaud
LIAFA, Université Paris 7

February 21, 2000

Summary by Marianne Durand

Abstract

A group automaton is a complete deterministic automaton such that each letter of the alphabet acts on the set of states as a permutation [1, 5]. The aim is to describe an algorithm for the random generation of a minimal group automaton with n states. The treatment is largely based on properties of random permutations and random automata.

1. Properties

A group automaton is a complete deterministic automaton such that each letter of the alphabet acts on the set of states as a permutation [1, 5]. We consider a group automaton $\mathcal{A}$, with states $1, 2, \ldots, n$. The state 1 is the initial state; the set of final states is denoted by F, the alphabet by $a, b, \ldots$, and the transitions by $q_2 = \delta(q_1, a)$ or equivalently (q_1, a, q_2).

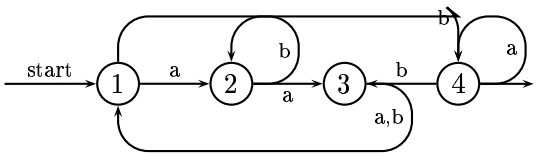

FIGURE 1. A group automaton.

Let us recall that two states q_1 and q_2 of an automaton are equivalent, notationally $q_1 \sim q_2$, if for every word u, the state $\delta(q_1, u)$ belongs to F if and only if $\delta(q_2, u)$ belongs to F. The automaton $\mathcal{A}$ is minimal if $\mathcal{A}$ has no distinct equivalent states. The structure properties of group automata are: the minimal automaton of a group automaton is a group automaton; the set of group automata is closed under union, intersection and complementation but it is not closed under star and product. As each letter acts like a permutation on the set of states, there cannot exist two transitions (q_1, a, q) and (q_2, a, q) with q_1 and q_2 distinct. This means that there is a "reversibility" property because when the automaton is in a state q after reading a word u, it is possible to retrace the path followed.

We are now interested in the connexity of an automaton. An automaton is connected if for any state q, there is a path joining the initial state to q. Because of the reversibility property, if a group automaton is connected then it is strongly connected, which means that for any states q and q', there is a path from q to q'. A group automaton is defined by the k permutations coding the transitions and by the set F, where k is the cardinality of the alphabet, so there are $(2^n - 1)n!^k$ group automata. We show that, if the alphabet has at least two letters, almost all group automata on n states are connected. In order to do this we first state the fact that given two permutations

σ and α the generated group $\langle \sigma, \alpha \rangle$ is almost surely transitive. This can be shown by a simple combinatorial argument. Take two letters a and b and consider σ_a the permutation related to a and σ_b the one related to b; then as $\langle \sigma_a, \sigma_b \rangle$ is almost always transitive the automaton is almost always connected. We even have an asymptotic estimate if the alphabet has exactly two letters:

$$\frac{\text{Card(not connected group automata)}}{\text{Card(group automata)}} \sim \frac{1}{n}.$$

2. Minimality

We now have to study the minimality of the automaton. An important theorem is that almost all connected group automata are minimal. The proof is partially based on the study of the one-letter case: if the automaton is connected, then as there is only one letter a, the permutation induced by a is a circular permutation. It is minimal if it is not stable under a rotation which is equivalent to saying that the word $u = 1 \cdots \delta^k(1,a) \cdots \delta^{n-1}(1,a)$ is not a non-trivial factor of uu. Then in this case by counting the words corresponding to minimal circular permutations we show that almost all connected automata are minimal on a one-letter alphabet. If the alphabet has more than one letter, we observe that for almost all group automata, there is a letter a such that the permutation induced by a on the set of states has only one cycle of maximum length [3]. More precisely, we have the following lemma:

Lemma 1. *The probability that a permutation σ of size n has more than two cycles of maximum length is $o(1)$.*

Proof. Let $c_{n,m}$ be the probability that a permutation of size n has exactly two maximal cycles of size $m+1$. We note the generating function $C_m(z) = \sum_{n=0}^{\infty} c_{n,m} z^n$ and $c_n = \sum_{m \le n/2} c_{n,m}$. The following equality holds:

$$C_m(z) = \frac{z^{2(m+1)}}{2(m+1)^2} e^z \cdots e^{\frac{z^m}{m}} = \frac{1}{1-z} \frac{z^{2(m+1)}}{2(m+1)^2} \exp\bigl(-r_m(z)\bigr)$$

where $r_m(z) = \sum_{n>m} z^n/n$ is the remainder of the generating function of the logarithm. In order to get the coefficient $c_{n,m}$ we apply Cauchy's formula:

$$c_{n,m} = \frac{1}{2i\pi} \int_{\mathcal{C}} \frac{1}{1-z} \frac{z^{2(m+1)}}{2(m+1)^2} \exp\bigl(-r_m(z)\bigr) \frac{dz}{z^{n+1}}$$

where $\mathcal{C}$ is a path around the origin. We choose for this path a circle around the origin defined by: $|z| = e^{-1/n}$ and we set $z = e^{-p/n}$ for a change of variable. So we have

$$c_{n,m} = \frac{1}{2i\pi} \int_{1-in\pi}^{1+in\pi} \frac{\exp\bigl(-r_m(e^{-p/n})\bigr)}{1-e^{-p/n}} \frac{e^{-p(2m+2)/n}}{2(m+1)^2} \frac{e^p}{n}\, dp$$

We now need to approximate some of the quantities in the integral, for this we use a technique and a few lemmas provided in [2]. We first have the relations

$$r_m(e^{-p}) = E(mp) + O\left(\frac{e^{-mp}}{m}\right) \quad \text{and} \quad \frac{1}{n\left(1-e^{-p/n}\right)} = \frac{1}{p} + \frac{1}{n}\psi\left(\frac{p}{n}\right) \tag{1}$$

with $E(x) = \int_x^{\infty} \frac{e^{-v}}{v}\, dv$ and $\psi(z) = \frac{1}{1-e^{-z}} - \frac{1}{z}$, and where the error term $O\bigl(\exp(-mp)/M\bigr)$ is moreover uniform over $\Re(p) > 0$ and $\bigl|\Im(p)\bigr| \le \pi$.

Property 1. *For all $a > 0$, the function $e^{-aE(u)}$ is bounded on $\Re(u) > 0$.*

The relations 1 allow us to write, after we set $\mu = m/n$:

$$c_{n,m} = \frac{1}{2i\pi}\int_{1-in\pi}^{1+in\pi} \exp\left(-E(\mu p) + O\left(\frac{1}{m}\right)\right)\left(\frac{1}{p} + \frac{1}{n}\psi\left(\frac{p}{n}\right)\right)\frac{e^p e^{-p(2m+2)/n}}{2(m+1)^2}\,dp$$
$$= \frac{1}{2i\pi}\int_{1-in\pi}^{1+in\pi} \exp\big(-E(\mu p)\big)\left(\frac{1}{p} + \frac{1}{n}\psi\left(\frac{p}{n}\right) + O\left(\frac{1}{pm}\right)\right)\frac{e^p e^{-p(2m+2)/n}}{2(m+1)^2}\,dp.$$

This rewrites as $c_{n,m} = I_1 + I_2 + I_3$ where

$$I_1 = \frac{1}{2i\pi}\int_{1-in\pi}^{1+in\pi} \exp\big(-E(\mu p)\big)\frac{1}{p}\frac{e^p e^{-p(2m+2)/n}}{2(m+1)^2}\,dp,$$
$$I_2 = \frac{1}{2i\pi}\int_{1-in\pi}^{1+in\pi} \exp\big(-E(\mu p)\big)\frac{1}{n}\psi\left(\frac{p}{n}\right)\frac{e^p e^{-p(2m+2)/n}}{2(m+1)^2}\,dp,$$
$$I_3 = \frac{1}{m}\frac{1}{2i\pi}\int_{1-in\pi}^{1+in\pi} \exp\big(-E(\mu p)\big)O\left(\frac{1}{p}\right)\frac{e^p e^{-p(2m+2)/n}}{2(m+1)^2}\,dp.$$

To study these three expressions, we use the fact that the quantities $\exp\big(-E(\mu p)\big)$ (Property 1) and $e^p e^{-p(2m+2)/n}$ are bounded uniformly on m. This helps us to give an upper bound for these three expressions: first,

$$I_1 = \int_{1-in\pi}^{1+in\pi} \frac{O(1)}{pm^2}\,dp = O\left(\frac{\log n}{m^2}\right)$$

and this approximation is uniform on m. Second

$$I_2 = \int_{1-in\pi}^{1+in\pi} O(1)\frac{1}{n}\psi\left(\frac{p}{n}\right)\frac{1}{2(m+1)^2}\,dp$$

as ψ is also bounded uniformly on m we have

$$I_2 = \frac{1}{nm^2}\int_{1-in\pi}^{1+in\pi} O(1)\,dp = O\left(\frac{1}{m^2}\right).$$

Third, as in the case of I_1, we obtain

$$I_3 = \frac{1}{m}\int_{1-in\pi}^{1+in\pi} O\left(\frac{1}{p}\right)\frac{1}{2(m+1)^2}\,dp = O\left(\frac{\log n}{m^3}\right).$$

Combining these estimates we obtain $c_{n,m} = O\left(\frac{\log n}{m^2}\right)$ uniformly on m. The approximation is going to be useful when m is greater than $\sqrt{n}$; otherwise we use the following lemma:

Lemma 2. *The probability that a permutation σ of size n has a maximal cycle of length smaller than $\sqrt{n}$ is $o(1)$.*

Proof. Let $p_{n,m}$ be the probability that a permutation of size n has all its cycles of size smaller than m. The saddle-point method gives us an upper bound for the quantity $p_{n,m}$. Then we have

$$p_{n,m} = [z^n]e^{l_m(z)} \le \frac{e^{l_m(r)}}{r^n} \quad \text{where} \quad l_m(z) = z + \cdots + \frac{z^m}{m}.$$

The saddle-point method drives us to apply this inequality to the value $r = n^{\frac{1}{3m}}$ chosen to fit the minimum, which gives

$$p_{n,m} \le \frac{\exp\left(n^{1/3}\log m\right)}{n^{n/3m}}, \quad \text{so} \quad p_{n,\sqrt{n}} \le e^{\left(\frac{n^{1/3}}{2} - \frac{\sqrt{n}}{3}\right)\log n} = o(1).$$

□

The probability that a permutation has two maximal cycles of size m is bounded by the probability that a permutation has one maximal cycle of size m. Therefore the probability that a permutation of size n has two maximal cycles of size smaller than $\sqrt{n}$ is $o(1)$. So $c_n = o(1) + \sum_{m=\sqrt{n}}^{m=n/2} c_{n,m} = o(1)$ by the approximation $c_{n,m} = O\left(\frac{\log n}{m^2}\right)$. Lemma 1 directly follows by showing that almost all permutations of size n having at least two maximal cycles have exactly two maximal cycles. □

We define $\mathcal{E}_n$ as the set of group automata $\mathcal{A}$ of size n that are connected and with the property that there exists one letter a such that the permutation induced by a has only one maximal cycle. By Lemma 1, we show that almost all connected group automata belong to $\mathcal{E}_n$. Furthermore, if $\mathcal{A}$ belongs to $\mathcal{E}_n$ then we can show that the maximal cycle of σ_a does not interfere with other cycles, because of their different cardinalities and so we can use the one-letter case, and say that this maximal cycle is almost always minimal. As the automaton considered is connected, this implies that the automaton is minimal. So we have the following result:

Theorem 1. *Almost all group automata are minimal.*

Proof. $\mathcal{E}_n \subset \text{Minimal}_n \subset \text{Connected}_n \subset \text{Group Automaton}_n$, and we have proved that almost every group automaton is in $\mathcal{E}_n$. □

3. Algorithm

This work naturally leads to an algorithm for generating uniformly at random a minimal connected group automata. Here the cardinality of the alphabet is bounded. The size of an automaton is the number n of states of its minimal automaton. The algorithm is:

- generate a random group automaton $\mathcal{A}$ using a function returning a random permutation for each letter of the alphabet. The cost is $O(n)$;
- test if $\mathcal{A} \in \mathcal{E}_n$, if not use Hopcroft's algorithm to check if it is minimal. Since Hopcroft is used rarely, the cost is $O(n)$;

this being done a constant number of time on average, because of the theorem above.

This yields a linear complexity in the average case, which is better than the best known algorithm by Hopcroft [4] which has complexity $n \log n$.

Bibliography

[1] Eilenberg (Samuel). – *Automata, languages, and machines. Vol. A.* – Academic Press, New York, 1974, xvi+451p. Pure and Applied Mathematics, Vol. 58.

[2] Gourdon (Xavier). – *Combinatoire, algorithmique et géométrie des polynômes.* – Thèse, École polytechnique, 1996.

[3] Gourdon (Xavier). – Largest component in random combinatorial structures. In *Proceedings of the 7th Conference on Formal Power Series and Algebraic Combinatorics (Noisy-le-Grand, 1995)*, vol. 180, pp. 185–209. – 1998.

[4] Hopcroft (John). – An $n \log n$ algorithm for minimizing states in a finite automaton. In *Theory of machines and computations (Proc. Internat. Sympos., Technion, Haifa, 1971).* pp. 189–196. – Academic Press, New York, 1971.

[5] Hopcroft (John E.) and Ullman (Jeffrey D.). – *Introduction to automata theory, languages, and computation.* – Addison-Wesley Publishing Co., Reading, Mass., 1979, x+418p. Addison-Wesley Series in Computer Science.